CLAIMS:

1. A method for obtaining a candidate nucleotide sequence, the candidate nucleotide sequence being indicative of a sequence of a target polynucleotide molecule T, T producing a hybridization signal $I(\bar{x})$ upon incubating T with a polynucleotide $\bar{x}$ for each polynucleotide $\bar{x}$ in a set E of polynucleotides, the method comprising the steps of:

(a) for each polynucleotide $\bar{x}$ in the set E of polynucleotides, obtaining a probability $P_0(\bar{x})$ of the hybridization signal $I(\bar{x})$ when the sequence $\bar{x}$ is not complementary to a subsequence of T and a probability $P_1(\bar{x})$ of the hybridization signal when the sequence $\bar{x}$ is complementary to a subsequence of T; so as to obtain a probabilistic spectrum (PS) of T;

(b) assigning a score to each of a plurality of candidate nucleotide sequences, the score being based upon the probabilistic spectrum and upon at least one reference nucleotide sequence ~~H; and~~

(c) selecting one or more candidate nucleotide sequences having an essentially maximal score.

2. The method according to Claim 1, wherein the polynucleotides $\bar{x}$ in the set E are immobilized on a surface.

3. The method according to Claim 1 ~~or 2,~~ wherein the set E is a set of k-mers.

4. The method according to Claim 3 wherein E is the set of all k-mers formed from nucleotides from a predetermined set of nucleotides..

5. The method of Claim 4 wherein the predetermined set of nucleotides is selected from the group consisting of

(a) adenine, guanine, cytosine, and thymine; and

(b) adenine, guanine, cytosine, uracil.

6. The method according to ~~any one of the previous claims,~~ wherein the score of a candidate nucleotide sequence $\hat{T}$ is based upon $L^e(\hat{T})$ where

$$L^e(\hat{T}) = \prod_{\bar{x} \in A} P_{\hat{T}(\bar{x})}(\bar{x}),$$

wherein $\bar{T}(\vec{x}) = 0$ if the sequence of $\vec{x}$ is not complementary to a subsequence of $\hat{T}$ and $\hat{T}(x) = 1$ if the sequence of $\vec{x}$ is complementary to a subsequence of $\hat{T}$.

7. The method according to ~~any one of Claims 1 to 6~~, wherein the score of a candidate sequence $\hat{T}$ is based upon $\tilde{L}^e(\hat{T})$ where $\log \tilde{L}^e(\hat{T}) = \sum_{i=0}^{m} \omega(e_i)$, wherein $\hat{T}$ contains polynucleotides $e_o, \ldots e_m$ and $\omega(e_i) = \log \frac{P_1(e_i)}{P_0(e_i)}$.

8. The method according to ~~any one of the previous claims~~, wherein T and H have a common length.

9. The method according to Claim 8, wherein the score of a candidate sequence $\hat{T}$ is based upon $D^u(\hat{T})$ where $D^u(\hat{T}) = \prod_{j=1}^{l} M^{(j)}[t_j, h_j]$, wherein $M^{(j)}[t_j, h_j]$ is a probability of a nucleotide t_j in position j of T being replaced with nucleotide h_j in position j of H.

10. The method according to Claim 9, wherein the score of a candidate nucleotide sequence $\hat{T}$ is $Score^u{}_1(\hat{T})$, or $Score^u{}_2(\hat{T})$ where

$Score^u{}_1(\hat{T}) = \log L^e(\hat{T}) + \log D^u(\hat{T})$ *and* $Score^u{}_2(\hat{T}) = \log \tilde{L}^e(\hat{T}) + \log D^u(\hat{T})$.

11. The method according to Claim 10 wherein the polynucleotides in the set E are k-mers and the step of selecting a candidate nucleotide sequence having an essentially maximal score comprises the steps of

(a) For each (k-1)-mer $\vec{y}$ calculating $S^u[\vec{y}, k-1] = \sum_{j=1}^{k-1} L^{(j)}[y_j, h_j$

(b) for each integer j = k, ...l,

(ba) for each polynucleotide sequence $(y_1, \ldots y_{k-1})$

(baa) calculating

$$S^u[\vec{y}, j] = L^{(j)}[y_{k-1}, h_j] + \max_{e=(\vec{z},\vec{y})\in E} \{S^u[z, j-1] + \omega(e)\}$$

wherein $L^{(j)}[y, h_j] = log\, M^{(j)}[y, h_j]$.

(bab) selecting a (k-1)-mer $P[\vec{y}, j]$

satisfying

$$S^u[P[\vec{y},j],j-1]+\omega((P[\vec{y},j],\vec{y}))=\max_{e=(\vec{z},\vec{y})\in E}\{S^u[z,j-1]+\omega(e)\}$$

(c) selecting a (k-1)-mer Z^l having a score essentially equal to $\max_{\vec{y}\in V} S^u[\vec{y},l]$;

(d) for j=k-1,...,l-1; recursively calculating (k-1)-mers Z^j where $Z^{j-1}=P[Z^j,j]$

(e) selecting candidate target sequence $<z^{k-1}{}_1, z^{k-1}{}_2,\ldots z^{k-1}{}_{k-1}, z^{k}{}_{k-1}, z^{k+1}{}_{k-1},\ldots z^{l}{}_{k-1}>$, where $Z^j=<z^j{}_1 z^j{}_2\ldots z^j{}_{k-1}>$

12. The method according to Claim 9, wherein the polynucleotides in the set E are k-mers, and the step of selecting a candidate nucleotide sequence having an essentially maximal score comprises the steps of:

(a) If the length *l* of the target is greater than the predetermined constant, settting $m=\frac{l+k-1}{2}$;

(b) For each $j = k-1, \ldots, m$, computing $S^u[\vec{y},j]$ according to Claim 10 for all $\vec{y}$;.

(c) For each $j = l, l-1, \ldots, m$, computing $R^u[\vec{y},j]$ according to equations (14) and (15) for all $\vec{y}$;

(d) Selecting $\vec{y}_m = \arg\max_{\vec{y}\in V}\{S^u[\vec{y},m]+R^u[\vec{y},m]$;

(e) Computing the optimal sequence aligned to $\langle h_1\ldots h_m\rangle$ ending with $\vec{y}_m$, and the optimal sequence aligned to $\langle h_m\ldots h_l\rangle$ beginning with $\vec{y}_m$.

13. The method according to ~~any one of Claims 1 to 7,~~ claim 1 wherein H and T have lengths such that the length of T is less than the length of H.

14. The method according to Claim 13, wherein the step of assigning a score to each of a plurality of candidate nucleotide sequences and the step of selecting the candidate target sequence are performed according to Algorithm B.

15. The method according to ~~any one of Claims 1 to 7~~, wherein H and T have arbitrary lengths.

16. The method according to Claim 15, wherein the step of assigning a score to each of a plurality of candidate nucleotide sequences and the step of selecting the candidate target sequence are performed according to Algorithm C.

17. The method according to Claim 15, wherein the step of assigning a score to each of a plurality of candidate nucleotide sequences and the step of selecting the candidate target sequence are performed according to Algorithm D.

18. The method according to Claim 17 wherein a Hidden Markov Model is used instead of a reference sequence.

19. The method according to ~~any one of Claims 1 to 11~~, wherein the algebraic equation (12a') replaces the algebraic equation (12a), the algebraic equation (12b') replaces the algebraic equation (12b), the algebraic equation (15') replaces the algebraic equation (15), and the algebraic equation (16') replaces the algebraic equation (16).

20. The method according to ~~any one of Claims 13,14, or 19~~, wherein the algebraic equation (20') replaces the algebraic equation (20), and the algebraic equation (21') replaces the algebraic equation (21).

21. The method according to ~~any one of Claims 15, 17, or 18~~, wherein the algebraic equation (29') replaces the algebraic equation (29), and the algebraic equation (30') replaces the algebraic equation (30).

22. The method according to ~~any one of the previous claims~~ wherein the target comprises two or more polynucleotide molecules.

23. The method according to ~~any one of Claims 1 to 22~~ computing the exact score $L^e(\hat{T})$ for several candidate sequences chosen according to the value of the approximated score $\tilde{L}^e(\hat{T})$.

24. The method according to ~~any one of the previous claims~~ further comprising a step of deleting candidate sequences having likelihood below a predetermined value.

25. The method according to ~~any one of Claims 1 to 5, 8 to 24~~, wherein the score of a candidate nucleotide sequence $\hat{T}$ is based upon $\underline{L}^e(\hat{T})$ where

$$\underline{L}^e(\hat{T}) = \prod_{\vec{x} \in A} P_{\underline{T}(\vec{x})}(\vec{x}),$$

wherein $\underline{T}(\vec{x}) = r$ if the sequence of $\vec{x}$ is complementary to exactly r subsequences of $\hat{T}$.

26. The method according to ~~any one of the previous claims~~, wherein the set E of polynucleotide does not include all the polynucleotide of a specific length.

27. The method according to ~~any one of the previous claims~~, wherein the set E of polynucleotide includes polynucleotides of different lengths.

28. The method according to ~~any one of the previous claims~~ for use in a task selected from the group comprising:

(a) Detecting or genotyping of Single Nucleotide Polymorphisms.

(b) Detecting or genotyping of genetic syndroms or disorders.

(c) Detecting or genotyping somatic mutations.

(d) Sequencing a polynucleotide whose function is related to the function of the reference polynucleotide.

29. The method according to ~~any one of the previous claims~~, wherein polynucleotides contain gaps, or universal bases.

30. The method according to ~~any one of the previous claims~~, wherein polypeptides are sequenced instead of polynucleotides.

31. A program storage device readable by machine, tangibly embodying a program of instructions executable by the machine to perform method steps for obtaining a candidate nucleotide sequence, the candidate nucleotide sequence being indicative of a sequence of a target polynucleotide molecule T, T producing a hybridization signal $I(\vec{x})$ upon incubating T with a polynucleotide $\vec{x}$ for each polynucleotide $\vec{x}$ in a set E of polynucleotides, the method comprising the steps of:

(a) for each polynucleotide $\vec{x}$ in the set E of polynucleotides, obtaining a probability $P_0(\vec{x})$ of $I(\vec{x})$ when the sequence $\vec{x}$ is not complementary to a subsequence of T and a probability $P_1(\vec{x})$ of $I(\vec{x})$ when the sequence $\vec{x}$ is

complementary to a subsequence of T; so as to obtain a probabilistic spectrum (PS) of T;

(b) assigning a score to each of a plurality of candidate nucleotide sequences, the score being based upon the probabilistic spectrum and upon at least one reference nucleotide sequence H; and

(c) selecting a candidate nucleotide sequence having an essentially maximal score.

32. A computer program product comprising a computer useable medium having computer readable program code embodied therein for obtaining a candidate nucleotide sequence, the candidate nucleotide sequence being indicative of a sequence of a target polynucleotide molecule T, T producing a hybridization signal $I(\vec{x})$ upon incubating T with a polynucleotide $\vec{x}$ for each polynucleotide $\vec{x}$ in a set E of polynucleotides, the computer program product comprising:

(a) for each polynucleotide $\vec{x}$ in the set E of polynucleotides, computer readable program code for causing the computer to obtain a probability $P_0(\vec{x})$ of $I(\vec{x})$ the sequence $\vec{x}$ is not complementary to a subsequence of T and a probability $P_1(\vec{x})$ of $I(\vec{x})$ when the sequence $\vec{x}$ is complementary to a subsequence of T;

(b) computer readable program code for causing the computer to assign a score to each of a plurality of candidate nucleotide sequences, the score being based upon the probabilistic spectrum and upon at least one reference nucleotide sequence H; and

(c) computer readable program code for causing the computer to select a candidate nucleotide sequence having an essentially maximal score.

ADD B27